MEASUREMENT SCIENCE REVIEW, **16**, (2016), No. 4, 211–217

DE GRUYTER OPEN

ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: http://www.degruyter.com/view/j/msr

The Cosine Error: A Bayesian Procedure for Treating a Non-repetitive Systematic Effect

Ignacio Lira[1] and Dieter Grientschnig[2]

[1] *Department of Mechanical and Metallurgical Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile, ilira@ing.puc.cl*
[2] *Metrolytik, Kernstockstr. 4, 8600 Bruck an der Mur, Austria*

An inconsistency with respect to variable transformations in our previous treatment of the cosine error example with repositioning (*Metrologia*, vol. 47, pp. R1–R14) is pointed out. The problem refers to the measurement of the vertical height of a column of liquid in a manometer. A systematic effect arises because of the possible deviation of the measurement axis from the vertical, which may be different each time the measurement is taken. A revised procedure for treating this problem is proposed; it consists in straightforward application of Bayesian statistics using a conditional reference prior with partial information. In most practical applications, the numerical differences between the two procedures will be negligible, so the interest of the revised one is mainly of conceptual nature. Nevertheless, similar measurement models may appear in other contexts, for example, in intercomparisons, so the present investigation may serve as a warning to analysts against applying the same methodology we used in our original approach to the present problem.

Keywords: Cosine error, probability density function, Bayesian statistics, reference prior, Borel's paradox.

1. Introduction

One of the examples appearing in a previous paper of ours [1] was the measurement of the vertical height of a column of liquid in a manometer, designated as Z. Our model was

$$Z = X\,F, \tag{1}$$

where X is the height indicated by the instrument and F is a correction factor for a systematic effect caused by a possible deviation of the measurement axis from the vertical. This correction factor was in turn modelled as

$$F = \cos Y, \tag{2}$$

where for Y – the misalignment angle – prior information about the range of its possible values was supposed to be available.

In a first instance, we assumed that indications $\boldsymbol{x} = \{x_1,\dots,x_n\}$ for X were obtained while keeping the instrument in a fixed position and that the angle Y was equally likely of being in the interval between given values $-y_1$ and y_2, where y_1 and y_2 are both positive and different in general. While formulae for obtaining the best estimate of the measurand Z and its associated standard uncertainty under these conditions are given in the Guide to the Expression of Uncertainty in Measurement (GUM) [2, subclause F.2.4.4], in [1] we complemented that analysis by deriving the probability density function (PDF) for Z using the Bayesian methodology.

In a second instance, we assumed – more realistically – that the data $\boldsymbol{x}$ were obtained by repositioning the device after every observation while keeping it in a fixed vertical plane. In this case, the measurement model changes to

$$Z = X_i F_i, \qquad i = 1,\dots,n, \tag{3}$$

where $F_i = \cos Y_i$ and the X_i's are the measured heights corresponding to *different* unknown misalignment angles Y_i, which can again be supposed to assume any value within the interval $[-y_1, y_2]$ with equal probability. The data are still the set $\boldsymbol{x}$, but they are now interpreted as a single datum for each of the quantities $\{X_1,\dots,X_n\}$, which we designate by $\boldsymbol{X}$.

Unfortunately, an inconsistency in our solution to this alternative formulation of the problem went unnoticed. In Section 2 of the present paper, a summary of that solution is given and the origin of the inconsistency is pointed out. In Section 3, we present an approach that is free from this shortcoming. Examples and discussion follow in Section 4 and conclusions are given in Section 5.

2. The original procedure

In our previous paper, we denoted by $f_A(\alpha\,|\,\mathscr{K})$ the PDF for a generic quantity A with possible values α, where $\mathscr{K}$ represents the given model, data, assumptions and other information. Here we shall keep the same notation, using ζ, ξ_i, ϕ_i and

DOI: 10.1515/msr-2016-0026

η_i for the possible values of quantities Z, X_i, F_i and Y_i, respectively. However, in [1] we used the delta function approach for effecting variable transformations; in the present paper we shall use instead the change-of-variables theorem [3, 4]. For simplicity, we shall analyse only the symmetrical case and take both angles y_1 and y_2 as equal to a given value y. Thus, our goal is to derive an expression for $f_Z(\zeta\,|\,\boldsymbol{x},y)$.

The original procedure started by considering each of the n models (3) at a time. The joint PDFs for the input quantities to these models are of the form

$$f_{X_i,F_i}(\xi_i,\phi_i\,|\,x_i,y) = f_{X_i}(\xi_i\,|\,x_i)f_{F_i}(\phi_i\,|\,y), \tag{4}$$

where we used the fact that X_i and F_i are independent *a priori*. (This is because, before knowing Z, having only information about X_i does not yield any information about F_i, and vice versa.) Change of variables produces

$$f_{X_i,Z}(\xi_i,\zeta\,|\,x_i,y) = \xi_i^{-1}f_{X_i}(\xi_i\,|\,x_i)\,f_{F_i}\left(\zeta\,\xi_i^{-1}\,|\,y\right) \tag{5}$$

and multiplication of all these PDFs gives

$$f_{\boldsymbol{X},Z}(\boldsymbol{\xi},\zeta\,|\,\boldsymbol{x},y) \propto f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})\prod \xi_i^{-1}f_{F_i}\left(\zeta\,\xi_i^{-1}\,|\,y\right), \tag{6}$$

where

$$f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x}) = \prod f_{X_i}(\xi_i\,|\,x_i) \tag{7}$$

with both products running from 1 to n. (The same convention holds for all other products and summations below.)

We then proceeded to derive expressions for PDFs $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and $f_{F_i}(\phi_i\,|\,y)$. The former was easily obtained by assuming that the indications x_i are drawn from Gaussian distributions centered at those readings and with a common unknown standard deviation S whose possible values σ range from zero to infinity. The individual likelihoods must then include this nuisance parameter, so they become

$$\ell(\xi_i,\sigma\,;x_i) \propto \sigma^{-1}\exp\left(-0.5\,\sigma^{-2}(\xi_i-x_i)^2\right) \tag{8}$$

and therefore the joint likelihood is

$$\ell(\boldsymbol{\xi},\sigma;\boldsymbol{x}) \propto \sigma^{-n}\exp\left(-0.5\,\sigma^{-2}\Gamma_X\right), \tag{9}$$

where

$$\Gamma_X = \sum(\xi_i-x_i)^2. \tag{10}$$

The quantities $\boldsymbol{X}$ in this likelihood are of equal inferential interest, which is higher than that of S. As is well known, a reasonable non-informative prior in these circumstances is the two-group reference prior [5], where one group is made up of the location parameters $\boldsymbol{\xi}$ and the other contains only the scale parameter σ. For the statistical model (9), this prior reads $f^o_{\boldsymbol{X},S}(\boldsymbol{\xi},\sigma) \propto \sigma^{-1}$. Then, according to Bayes' theorem, the joint posterior becomes

$$f_{\boldsymbol{X},S}(\boldsymbol{\xi},\sigma\,|\,\boldsymbol{x}) \propto \sigma^{-(n+1)}\exp\left(-0.5\,\sigma^{-2}\Gamma_X\right), \tag{11}$$

from which the parameter σ can be integrated out. The result is

$$f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x}) \propto \Gamma_X^{-n/2}. \tag{12}$$

To derive the PDFs $f_{F_i}(\phi_i\,|\,y)$, we use the functions $f_{Y_i}(\eta_i\,|\,y)$ (which are uniform with support on the interval $[-y,y]$) and change variables again. This gives

$$f_{F_i}(\phi_i\,|\,y) = \left(1-\phi_i^2\right)^{-1/2}f_{Y_i}(\pm\arccos\phi_i\,|\,y), \tag{13}$$

or equivalently,

$$f_{F_i}(\phi_i\,|\,y) \propto \left(1-\phi_i^2\right)^{-1/2} \qquad \text{for } \cos y \le \phi_i \le 1. \tag{14}$$

Substitution of (12) and (14) into (6) produces

$$f_{\boldsymbol{X},Z}(\boldsymbol{\xi},\zeta\,|\,\boldsymbol{x},y) \propto \Gamma_X^{-n/2}\prod\left(\xi_i^2-\zeta^2\right)^{-1/2}, \tag{15}$$

from which the desired PDF $f_Z(\zeta\,|\,\boldsymbol{x},y)$ can be obtained by marginalization followed by normalization.

Equivalently, transformation of (15) into the $(\boldsymbol{F},Z)$ and $(\boldsymbol{Y},Z)$ parameterizations gives, respectively,

$$f_{\boldsymbol{F},Z}(\boldsymbol{\phi},\zeta\,|\,\boldsymbol{x},y) \propto \Gamma_F^{-n/2}\prod\phi_i^{-1}\left(1-\phi_i^2\right)^{-1/2} \tag{16}$$

and

$$f_{\boldsymbol{Y},Z}(\boldsymbol{\eta},\zeta\,|\,\boldsymbol{x},y) \propto \Gamma_Y^{-n/2}\prod(\cos\eta_i)^{-1}, \tag{17}$$

where

$$\Gamma_F = \sum\left(\zeta\,\phi_i^{-1}-x_i\right)^2 \tag{18}$$

and

$$\Gamma_Y = \sum\left(\frac{\zeta}{\cos\eta_i}-x_i\right)^2. \tag{19}$$

Note that (17) corresponds with equation (45) in [1].

Is there anything wrong with this analysis? The answer is yes, but the reason is subtle. Basically, it is perfectly legitimate to multiply the PDFs $f_{X_i}(\xi_i\,|\,x_i)$ to give $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and to multiply the PDFs $f_{F_i}(\phi_i\,|\,y)$ to give $f_{\boldsymbol{F}}(\boldsymbol{\phi}\,|\,y)$. This is because, in the case at hand, the quantities $\boldsymbol{X}$ and $\boldsymbol{F}$ are all pairwise independent *a priori*, that is, before the measurement models (3) have been considered. Multiplying $f_{\boldsymbol{X}}(\boldsymbol{\xi}\,|\,\boldsymbol{x})$ and $f_{\boldsymbol{F}}(\boldsymbol{\phi}\,|\,y)$ to give $f_{\boldsymbol{X},\boldsymbol{F}}(\boldsymbol{\xi},\boldsymbol{\phi}\,|\,\boldsymbol{x},y)$ is also legitimate. But the latter PDF would not allow deriving the state-of-knowledge distribution about all input quantities *a posteriori*. If used in that sense, an inconsistency known as "Borel's paradox" would arise.

In the statistical literature, the circumstances giving rise to Borel's paradox have been discussed in [6–11]. Benefiting from those discussions, in [12] and [13] we showed that, in the metrological context, the effects of Borel's paradox are mainly revealed if further transformations of the quantities involved, designed as consistency checks, are carried out. For this reason, those inconsistencies were not noticed in our original analysis of the cosine error problem.

In our case, avoiding this paradox precludes a distribution of dimension lower than $2n$ to be derived uniquely from $f_{\boldsymbol{X},\boldsymbol{F}}(\boldsymbol{\xi},\boldsymbol{\phi}\,|\,\boldsymbol{x},y)$, as e.g. PDFs (15), (16) and (17), whose dimension is $n+1$. More precisely, the operation above that causes the trouble is the step from (5) to (6). This represents the multiplication of PDFs that share the common variable ζ. As discussed in Appendix 1, such an operation is ad hoc and leads to ambiguous results.

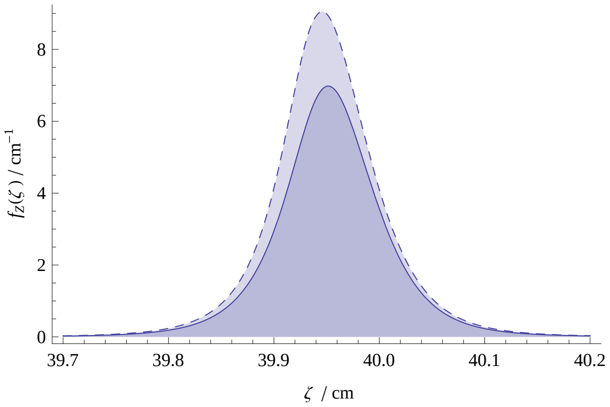

Fig. 1: PDFs for the measurand Z given the data set $\boldsymbol{x}_a$ with a maximum misalignment angle set to $y = 5°$. The solid line corresponds to figure 2b in [1]. The dashed line was obtained with an improved numerical recalculation of the original procedure; it virtually coincides with the PDF that we got with the revised procedure.

3. The revised procedure

Which would then be the proper way of deriving the posterior for the output quantity Z? To avoid Borel's paradox we will use neither (6) nor (11). Instead, we will express the likelihood (9) in the form that results from taking the models (3) into consideration. Because of the transformation invariance of the likelihood function – proved e.g. in [14, p. 99] – this is simply

$$\ell(\boldsymbol{\phi},\zeta,\sigma;\boldsymbol{x}) \propto \sigma^{-n}\exp\left(-0.5\,\sigma^{-2}\Gamma_F\right). \tag{20}$$

We now need a prior for $\boldsymbol{F}$, Z and σ. This can be written as

$$f^o_{\boldsymbol{F},Z,S}(\boldsymbol{\phi},\zeta,\sigma\,|\,y) = f^o_{Z,S|\boldsymbol{F}}(\zeta,\sigma\,|\,\boldsymbol{\phi})f_{\boldsymbol{F}}(\boldsymbol{\phi}\,|\,y), \tag{21}$$

where $f^o_{Z,S|\boldsymbol{F}}(\zeta,\sigma\,|\,\boldsymbol{\phi})$ is a reference prior conditional on the partial information represented by

$$f_{\boldsymbol{F}}(\boldsymbol{\phi}\,|\,y) \propto \prod\left(1-\phi_i^2\right)^{-1/2}. \tag{22}$$

In Appendix 2 we show the derivation of $f^o_{Z,S|\boldsymbol{F}}(\zeta,\sigma\,|\,\boldsymbol{\phi})$, which was performed along the lines of [15] and [16]. The result is

$$f^o_{Z,S|\boldsymbol{F}}(\zeta,\sigma\,|\,\boldsymbol{\phi}) \propto (\sigma\,\mu)^{-1}, \tag{23}$$

where

$$\mu = \min(\phi_1,\ldots,\phi_n). \tag{24}$$

From Bayes' theorem, the joint posterior is then

$$f_{\boldsymbol{F},Z,S}(\boldsymbol{\phi},\zeta,\sigma\,|\,\boldsymbol{x},y) \propto \ell(\boldsymbol{\phi},\zeta,\sigma;\boldsymbol{x})f_{\boldsymbol{F}}(\boldsymbol{\phi}\,|\,y)f^o_{Z,S|\boldsymbol{F}}(\zeta,\sigma\,|\,\boldsymbol{\phi}).$$

Substituting (20), (22) and (23) into (3), and integrating out the variable σ, we obtain

$$f_{\boldsymbol{F},Z}(\boldsymbol{\phi},\zeta\,|\,\boldsymbol{x},y) \propto \mu^{-1}\Gamma_F^{-n/2}\prod\left(1-\phi_i^2\right)^{-1/2} \tag{25}$$

or, in the $(\boldsymbol{Y},Z)$ parameterization,

$$f_{\boldsymbol{Y},Z}(\boldsymbol{\eta},\zeta\,|\,\boldsymbol{x},y) \propto \Gamma_Y^{-n/2}\left[\min(\cos\eta_1,\ldots,\cos\eta_n)\right]^{-1}. \tag{26}$$

4. Examples and discussion

4.1. The example in our previous paper

The simulated data we used in our previous paper was $\boldsymbol{x}_a$ / cm $= \{39.88, 39.93, 40.00, 40.09, 40.12\}$, with a maximum misalignment angle set to $y = 5°$. The PDF $f_Z(\zeta\,|\,\boldsymbol{x}_a, 5°)$ that we obtained using the original procedure was shown in figure 2b of [1] and is here reproduced by the solid line in Fig. 1. However, in that article we wrote a disclaimer stating that the numerical accuracy of our calculations was doubtful, because at that time we used the approximate quasi-Monte Carlo numerical integration method [17] as implemented in the Mathematica software. In the present research we used instead the more accurate global adaptive strategy [18] in the same software, and adjusted some of the parameters of the integration routine to avoid the convergence errors we had encountered previously. The result was the dashed line in Fig. 1, which we believe to be a more accurate representation of the actual shape of the PDF corresponding to the original procedure.

We then proceeded to marginalize $f_{\boldsymbol{Y},Z}(\boldsymbol{\eta},\zeta\,|\,\boldsymbol{x}_a,5°)$, given by (26), followed by normalization. The PDF that resulted was virtually indistinguishable from the dashed line in Fig. 1. Therefore, this dashed line replaces figure 2b in [1] and represents the PDF $f_Z(\zeta\,|\,\boldsymbol{x}_a,5°)$ obtained with both, the original and revised procedures. Its mean is 39.951 cm and its standard deviation is 0.060 cm.

Why is the difference between the two procedures so insignificant? By comparing equations (17) and (26), one sees that in the former, the factor $\Gamma_Y^{-n/2}$ is divided by the product of the cosines of the variables representing the unknown misalignment angles, while in the latter this same factor is divided solely by the minimum of these cosines. For a maximum misalignment angle of 5°, the mean of the product of the five cosines is 0.9937, whereas the mean of the cosine of the maximum of the five angles is 0.9973. That is the reason for the difference between the two procedures being numerically negligible in this case.

4.2. The effect of the maximum misalignment angle

As a second example, consider another data set in which the observations deviate from 40 cm by approximately half as much as in the previous example: $\boldsymbol{x}_b$ / cm $= \{39.94, 39.97, 40.00, 40.05, 40.06\}$. Because the mean of the maximum misalignment angle in this second example is even less than in the first, it was reasonable to find that, again, there are no appreciable differences between the two procedures. Fig. 2 shows the PDFs for Z corresponding to $y = 4°$ (dashed line) and $y = 5°$ (solid line). For clarity, the same PDFs are shown in Fig. 3, but in logarithmic ordinates. For $y = 4°$ the mean is 39.969 cm and the standard deviation is 0.028 cm. For $y = 5°$, the mean is 39.934 cm and the standard deviation is 0.015 cm.

It was surprising to find that these summary characteristics and the shapes of the two PDFs differ radically. This drastic

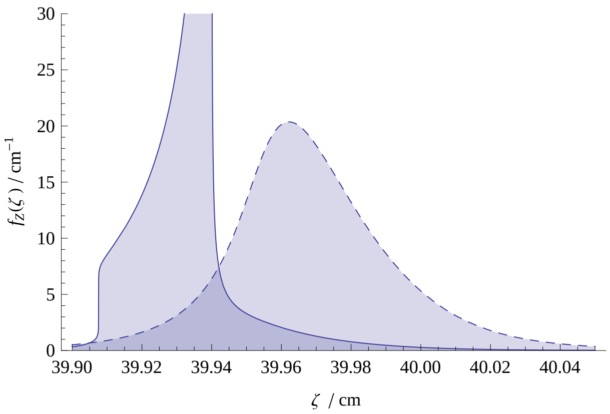

Fig. 2: For both the original and revised procedures, PDFs $f_Z(\zeta\,|\,\boldsymbol{x}_b, 4^\circ)$ (dashed line) and $f_Z(\zeta\,|\,\boldsymbol{x}_b, 5^\circ)$ (solid line). For clarity, the latter is shown clipped at an ordinate value of 30 cm^{-1}.

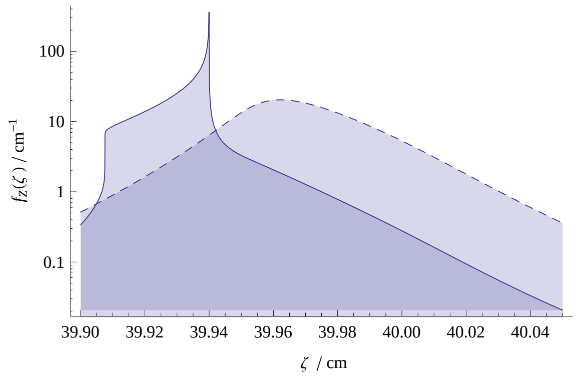

Fig. 3: Same as Fig. 2, except for logarithmic ordinate scale and clipping at 300 cm^{-1}. We found that the spike at $\zeta = 39.94$ cm rises to about 3 300 cm^{-1}.

change is due to the singularity of $f_{\boldsymbol{Y},Z}(\boldsymbol{\eta}, \zeta\,|\,\boldsymbol{x}, y)$ that occurs in both the original and revised procedures whenever

$$\zeta = x_1 \cos\eta_1 = \cdots = x_n \cos\eta_n. \qquad (27)$$

In Appendix 3 we provide details about the way we handled the occurrence of this condition. Evidently, it may only be fulfilled for values of ζ that are both less than or equal to $x_{\min}$ and larger than or equal to $x_{\max}\cos y$. Hence, such singularities may only arise if $x_{\max}\cos y \le x_{\min}$, i.e. if the maximum misalignment angle y is greater than the arc cosine of $x_{\min}/x_{\max}$, which may be designated as the critical angle y_c. Thus, no singularities arise in the first example ($y = 5^\circ$, $y_c = 6.27^\circ$) and for $y = 4^\circ$ in the second example ($y_c = 4.44^\circ$), in which case the corresponding PDFs are approximately bell-shaped. In contrast, by taking $y = 5^\circ$ in the second example, PDF $f_Z(\zeta\,|\,\boldsymbol{x}_b, 5^\circ)$ exhibits a peculiar central zone bounded by $\zeta = x_{\max}\cos y = 39.9076$ cm at the left and $\zeta = x_{\min} = 39.94$ cm at the right; it rises abruptly at the former bound and reaches a maximum at the latter, decreasing steeply thereafter.

Why does $f_Z(\zeta\,|\,\boldsymbol{x}_b, 5^\circ)$ exhibit such a sharp spike at $\zeta = x_{\min}$? Table 1 shows the cosine values $\phi_i = \cos\eta_i$ that fulfill condition (27) for $\zeta = 39.9076$ cm (second column) and $\zeta = 39.94$ cm (third column). It can be seen that the former series of values are farther away from unity than those for the latter. Consequently, the probability densities ensuing from (14) are smaller in the case of $\zeta = 39.9076$ cm. Furthermore, the abscissa value $\zeta = 39.94$ cm is the only one where condition (27) involves a cosine equal to 1 and thus, in addition to the singularity caused by (18), another singularity due to (14) becomes effective. The spike at $\zeta = 39.94$ cm arises because these two singularities coincide at that point.

Table 1: Cosine values that fulfil condition (27) for the two values that delimit the domain where singularities of the joint PDFs (17) and (26) arise. The second column corresponds to $\zeta = x_{\max}\cos y = 39.9076$ cm and the third column corresponds to $\zeta = x_{\min} = 39.94$ cm.

x_i / cm	ϕ_i	
39.94	0.99919	1.00000
39.97	0.99844	0.99925
40.00	0.99769	0.99850
40.05	0.99644	0.99725
40.06	0.99619	0.99700

A tentative physical explanation for the peculiar shape of $f_Z(\zeta\,|\,\boldsymbol{x}_b, 5^\circ)$ is that, for $x_{\max}\cos y < \zeta < x_{\min}$, it would be feasible (though perhaps not realistic) to explain the scatter of the observed data ranging from $x_{\min}$ to $x_{\max}$ solely by the occurrence of different misalignment angles within the interval $(-y, y)$. In contrast, if $\zeta < x_{\max}\cos y$ or $x_{\min} < \zeta$ the potential variations in misalignment are too small to cause the scatter observed and therefore the data themselves must be affected by random error. In other words, when $y > y_c$ the support of $f_Z(\zeta\,|\,\boldsymbol{x}, y)$ can be divided into two regions: a central portion where misalignment alone would theoretically suffice to explain the scatter of the data, delimited at both sides by portions where the data observed may not have come about without the occurrence of random measurement errors.

The implication of the behaviour just described is that, if the data were believed to be measured with negligible error, the maximum misalignment angle should be set to a value at least equal to y_c, otherwise the information available would be inconsistent. Thus, choosing $y = 5^\circ$ in the first example and $y = 4^\circ$ in the second would be unacceptable under the assumption of negligible random measurement errors, because these values are less than the corresponding critical angles ($y_c = 6.27^\circ$ and $y_c = 4.44^\circ$, respectively). However, the chosen maximum misalignment angles are plausible in view of the potential random errors (in these examples, the meniscus that forms and perhaps minor temperature variations might produce random measurement errors of the order of 0.3 cm). In the second example a value larger than y_c was assigned to y, so the singularity zone of the PDF to the left of the mode at $\zeta = x_{\min}$ dominates. The fact that in this zone observations devoid of random errors would theoretically be possible leads to a more focused PDF and therefore explains the counter-intuitive but nonetheless logical feature that a larger maximum misalignment angle (i.e. a less informative prior) should lead to a narrower posterior.

This is the only scenario known to us where a seemingly benign physical problem that features a Gaussian statistical model and simple rectangular input distributions results in a complex shape of the posterior with a needle-like peak. However, the scenario seems to be well behaved only when it is regarded in terms of angles since their informative priors are rectangular. By contrast, if not the angles but their cosines are considered, it becomes evident that in fact the situation is far from benign because the PDFs for the cosines have a singularity at 1. This causes the most striking feature of the marginal posterior, viz. the spike at $\zeta = x_{\min} = 39.94$ cm.

5. Conclusion

A deficiency in our previous treatment of the cosine error example with repositioning, given in [1], has been pointed out and a remedial procedure has been presented. The deficiency is due to our original analysis being affected by Borel's paradox, whose consequences are not immediately obvious but are nonetheless significant in that, as explained in Appendix 1, they might lead to contradicting outcomes should the posterior PDF for the measurand be subject to further transformations.

The revised procedure consists of using Bayesian statistics accompanied by a conditional reference prior derived in accordance with the principles in [19] for the case of partial information being previously available. In the examples presented, the numerical differences between the two procedures are negligible, so the interest of the revised one in the case at hand is only of conceptual nature. Nevertheless, the models (3) may appear in other contexts, for example, in intercomparison measurements. In that case, the quantities X_i would be those measured by the laboratories and the quantities F_i would be correction factors for systematic effects. The above investigation may serve as a warning to analysts against the shortcomings of a commonly used procedure of aggregating probability distributions by multiplication.

Acknowledgement

The financial support of the Chilean National Fund for Scientific and Technological Development (Fondecyt), research grant 1141165, is gratefully acknowledged.

Appendix 1: Aggregating probability distributions by multiplication

Aggregating PDFs by multiplication brings about ambiguities. To show this, consider just two measurement models of the form (3):

$$Z = X_1 F_1 = X_2 F_2. \tag{28}$$

Then, from (5),

$$\begin{aligned} f_{X_1,Z}(\xi_1,\zeta) &= \xi_1^{-1} f_{X_1}(\xi_1)\, f_{F_1}\left(\zeta\, \xi_1^{-1}\right) \\ f_{X_2,Z}(\xi_2,\zeta) &= \xi_2^{-1} f_{X_2}(\xi_2)\, f_{F_2}\left(\zeta\, \xi_2^{-1}\right), \end{aligned} \tag{29}$$

where we have dropped the 'given' part of the arguments of the functions to simplify notation.

Suppose we were really interested not in the height Z of the column of liquid in the manometer, but in some other quantity W related to the former by an arbitrary non-linear measurement model

$$Z = g(W). \tag{30}$$

To obtain the PDF $f_W(\omega)$ by multiplication of distributions we can proceed in either of two ways [12, Sect. 3]. The first consists in multiplying the two PDFs (29) and then transforming the result by means of (30). This procedure yields

$$\begin{aligned} f_{X_1,X_2,W}(\xi_1,\xi_2,\omega) &= \left|\frac{\mathrm{d}g(\omega)}{\mathrm{d}\omega}\right| \frac{1}{\xi_1\xi_2} \\ &\times f_{X_1}(\xi_1) f_{X_2}(\xi_2) f_{F_1}\left(\frac{g(\omega)}{\xi_1}\right) f_{F_2}\left(\frac{g(\omega)}{\xi_2}\right). \end{aligned} \tag{31}$$

The alternative method is to transform each of the PDFs (29) before multiplying them. In this way we get

$$\begin{aligned} f_{X_1,X_2,W}(\xi_1,\xi_2,\omega) &= \left(\frac{\mathrm{d}g(\omega)}{\mathrm{d}\omega}\right)^2 \frac{1}{\xi_1\xi_2} \\ &\times f_{X_1}(\xi_1) f_{X_2}(\xi_2) f_{F_1}\left(\frac{g(\omega)}{\xi_1}\right) f_{F_2}\left(\frac{g(\omega)}{\xi_2}\right). \end{aligned} \tag{32}$$

Thus, the discrepancy between these results is due to the Jacobian of the transformation entering once in the first approach but twice in the second.

To make eqs. (31) and (32) compatible, one may use the 'logarithmic pooling' technique, which consists in first raising each distribution to some exponent before multiplying them. It is easily seen that the inconsistency above disappears if the exponents are chosen such that they add up to one. We used this technique in a preliminary version of the present article [20]. However, logarithmic pooling is an ad hoc method, so we decided not to explore it further here.

Appendix 2: Derivation of reference prior (23)

With $\theta_1 = \zeta$ and $\theta_2 = \sigma$, the Fisher information matrix for the statistical model (9) becomes

$$I_{\theta\theta}(\boldsymbol{\phi},\boldsymbol{\theta}) = \mathrm{diag}\left(\theta_2^{-2}\sum\phi_i^{-2}, 2n\,\theta_2^{-2}\right), \tag{33}$$

(cf. (4) in [15]). Using the notation of [21], this implies

$$\left|I_{\theta\theta[\sim 00]}(\boldsymbol{\phi},\boldsymbol{\theta})\right| = 2n\,\theta_2^{-4}\sum\phi_i^{-2} \tag{34}$$

$$\left|I_{\theta\theta[\sim 11]}(\boldsymbol{\phi},\boldsymbol{\theta})\right| = 2n\,\theta_2^{-2} \tag{35}$$

$$\left|I_{\theta\theta[\sim 22]}(\boldsymbol{\phi},\boldsymbol{\theta})\right| = 1, \tag{36}$$

so Lemma 2.1 of [21] directly yields

$$|h_1(\boldsymbol{\theta})| = \theta_2^{-2}\sum\phi_i^{-2} \tag{37}$$

$$|h_2(\boldsymbol{\theta})| = 2n\,\theta_2^{-2}. \tag{38}$$

According to the notation of [22] these expressions factorize as

$$\left|h_j(\boldsymbol{\theta})\right| = h_{j1}\left(\boldsymbol{\theta}_{(j)}\right) h_{j2}\left(\boldsymbol{\theta}_{(j)}^C\right). \tag{39}$$

Only $h_{11}\left(\boldsymbol{\theta}_{(1)}\right)$ and $h_{21}\left(\boldsymbol{\theta}_{(2)}\right)$ are required to calculate with the help of Theorem 1 of [22] or Proposition 1 of [23] the conditional reference prior that results from compact rectangular subsets (i.e. subsets that are products of compact sets in the two subspaces whose bounds are independent). This prior is

$$p^*(\boldsymbol{\theta}|\boldsymbol{\phi}) = h_{11}^{1/2}\left(\boldsymbol{\theta}_{(1)}\right) h_{21}^{1/2}\left(\boldsymbol{\theta}_{(2)}\right). \tag{40}$$

It is plausible that the ϕ_i's have to be assigned to the factors that matter, because otherwise they could never take effect. However, this is just a supposition, not a reliable remedy for the ambiguity of the assignment. Since this calls the ensuing simplified calculation into question, it may be checked by a more cumbersome computation method referenced below. Nonetheless, in order to test the presumption made, let us proceed with the resulting factorization

$$h_{11}\left(\boldsymbol{\theta}_{(1)}\right) = \sum \phi_i^{-2}, \tag{41}$$

$$h_{12}\left(\boldsymbol{\theta}_{(1)}^C\right) = \theta_2^{-2}, \tag{42}$$

$$h_{21}\left(\boldsymbol{\theta}_{(2)}\right) = 2n\,\theta_2^{-2}, \tag{43}$$

$$h_{22}\left(\boldsymbol{\theta}_{(2)}^C\right) = 1. \tag{44}$$

This yields the conditional reference prior

$$p^*(\boldsymbol{\theta}|\boldsymbol{\phi}) = \theta_2^{-1}\left(\sum \phi_i^{-2}\right)^{1/2}. \tag{45}$$

By following the more fundamental procedure in [24] one obtains the same result, which retrospectively proves the above simplified derivation.

We continue in accordance with Section 3 of [16]. Since the quantities X, F and Z are all non-negative, the subsets of the original parameter space can be defined as e.g.

$$\Xi_{1,m} \times \ldots \times \Xi_{n,m} \times \Sigma_m = \{(\boldsymbol{\xi},\sigma) : \mathrm{e}^{-m} \le \xi_1 \le \mathrm{e}^m, \ldots, \mathrm{e}^{-m} \le \xi_n \le \mathrm{e}^m, \mathrm{e}^{-m} \le \sigma \le \mathrm{e}^m\}, \tag{46}$$

where $m = 1,2,\ldots$. (Note that in this expression Σ does not stand for a summation sign.) In order that none of these limits is exceeded, the subsets restricted to $\boldsymbol{\theta}$ take the form $\boldsymbol{\theta}_m = \{\boldsymbol{\theta} : L_{\inf} \le \theta_1 \le L_{\sup}, \mathrm{e}^{-m} \le \theta_2 \le \mathrm{e}^m\}$, where $L_{\inf} = \mathrm{e}^{-m}\max_{1\le l\le n}(\phi_l)$ and $L_{\sup} = \mathrm{e}^m \min_{1\le l\le n}(\phi_l)$. In analogy to (6) of [16] follows

$$K_m(\boldsymbol{\phi}) = \left[2m\left(\sum \phi_i^{-2}\right)^{1/2} \int_{L_{\inf}}^{L_{\sup}} \mathrm{d}\theta_1\right]^{-1}, \tag{47}$$

hence

$$\lim_{m\to\infty} \frac{K_m(\boldsymbol{\phi})}{K_m(\boldsymbol{\phi}^\times)} \propto \left[\min_{1\le l\le n}(\phi_l)\left(\sum \phi_i^{-2}\right)^{1/2}\right]^{-1}. \tag{48}$$

Thereby we get (23):

$$p(\boldsymbol{\theta}|\boldsymbol{\phi}) \propto \left[\theta_2 \times \min_{1\le l\le n}(\phi_l)\right]^{-1}. \tag{49}$$

Appendix 3: Integration details

To deal with the singularities of $f_{\boldsymbol{Y},Z}(\boldsymbol{\eta},\zeta\,|\,\boldsymbol{x},y)$, we tried several transformations. The one we found most useful is to introduce a new variable for

$$\left[\pm\left(\frac{\zeta}{\cos\eta_j} - x_j\right)\right]^{-4}, \tag{50}$$

where subscript j is such that x_j is an intermediate datum (in our examples we used $x_j = 40.00$ cm).

By this transformation we eliminated the strongest pole of the integrand. The integration range $(0,y)$ is thus split at the position of the pole and the latter is mapped to infinity. In this way, the integral on the finite support $(0,y)$ is replaced by the sum of two integrals whose supports both extend to infinity. The integration variables of these two integrals shall be denoted by ν or ω, respectively, according to whether the plus or minus sign in (50) applies.

This technique is only applicable if the term (50) has a pole within the interval $(0,y)$ for η_j. By using an intermediate datum for x_j, this is the case in an interval that extends at both sides farther than the domain where the troublesome singularities arise, which would not apply if the smallest or largest datum were used. Still, this technique is restricted to a limited portion of the support of the integral to be calculated, whereas outside of that portion integration of the original form of the integrand is employed.

With the transformation specified by (50), the joint PDF for $\boldsymbol{Y}$ and Z becomes

$$f_{V,W,\widetilde{\boldsymbol{Y}},Z}(\nu,\omega,\widetilde{\boldsymbol{\eta}},\zeta) = \left|\frac{\mathrm{d}g}{\mathrm{d}\nu}\right| f_{Y_j,\widetilde{\boldsymbol{Y}},Z}(g,\widetilde{\boldsymbol{\eta}},\zeta) + \left|\frac{\mathrm{d}h}{\mathrm{d}\omega}\right| f_{Y_j,\widetilde{\boldsymbol{Y}},Z}(h,\widetilde{\boldsymbol{\eta}},\zeta), \tag{51}$$

where, as in Appendix 1, we have again dropped the 'given' part of the arguments of the PDFs for conciseness. Vectors $\widetilde{\boldsymbol{Y}}$ and $\widetilde{\boldsymbol{\eta}}$ exclude quantity Y_j and its associated dummy variable η_j, respectively. The functions g and h are

$$g(\nu) = \arccos\left(\frac{\zeta}{x_j + \nu^{-1/4}}\right) \tag{52}$$

$$h(\omega) = \arccos\left(\frac{\zeta}{x_j - \omega^{-1/4}}\right). \tag{53}$$

For the original PDF (17), this gives

$$f_{V,W,\widetilde{\boldsymbol{Y}},Z}(\nu,\omega,\widetilde{\boldsymbol{\eta}},\zeta) \propto (\Lambda_V + \Lambda_W)\widetilde{\prod}(\cos\eta_i)^{-1}, \tag{54}$$

where

$$\Lambda_V = \left[4\nu^{5/4}\sqrt{\left(x_j+\nu^{-1/4}\right)^2-\zeta^2}\right]^{-1} \times \left[\nu^{-1/2}+\widetilde{\sum}\left(\frac{\zeta}{\cos\eta_i}-x_i\right)^2\right]^{-n/2} \quad (55)$$

$$\Lambda_W = \left[4\omega^{5/4}\sqrt{\left(x_j-\omega^{-1/4}\right)^2-\zeta^2}\right]^{-1} \times \left[\omega^{-1/2}+\widetilde{\sum}\left(\frac{\zeta}{\cos\eta_i}-x_i\right)^2\right]^{-n/2}. \quad (56)$$

In these equations, the product and summations are from 1 to n, but excluding j.

Transforming the revised PDF (26) gives similar expressions, except that Λ_V and Λ_W have to be multiplied by $\zeta/(x_j+\nu^{-1/4})$ and $\zeta/(x_j-\omega^{-1/4})$, respectively, and the product $\widetilde{\prod}(\cos\eta_i)^{-1}$ has to be substituted by the reciprocal cosine of the maximum of the deviation angles (substituting η_j by (52) or (53)). The integration over ν is from $(\zeta/\cos y - x_j)^{-4}$ to infinity and that over ω is from $(\zeta - x_j)^{-4}$ to infinity.

Because ordinary numerical integration resulted in artifacts or error messages, we computed $f_Z(\zeta\,|\,\boldsymbol{x}_b, 5^\circ)$ at a number of unequally spaced discrete points in the interval between $\zeta = 20$ cm and $\zeta = 60$ cm (with a higher concentration of points in the regions of rapid change of the PDF) and used a spline interpolating function to establish that posterior.

REFERENCES

[1] Lira, I., Grientschnig, D. (2010). Bayesian assessment of uncertainty in metrology: A tutorial. *Metrologia* 47, R1–R14.

[2] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM 1995 with minor corrections)*. JCGM 100:2008.

[3] Schwartz, J. (1954). The formula for change in variables in a multiple integral. *The American Mathematical Monthly*, 61 (2), 81–85.

[4] Lax, P. D. (1999). Change of variables in multiple integrals. *The American Mathematical Monthly*, 106 (6), 497–501.

[5] Fernández, C., Steel, M. F. J. (1999). Reference priors for the general location-scale model. *Statistics & Probability Letters*, 43 (4), 377–384.

[6] Rao, M. M. (1988). Paradoxes in conditional probability. *Journal of Multivariate Analysis*, 27 (2), 434–446.

[7] Wolpert, R. L. (1995). Comment (response to Raftery, A. E., Givens, G. H., Zeh, J. E.: Inference from a deterministic population dynamics model for bowhead whales.). *Journal of the American Statistical Association*, 90 (430), 426–427.

[8] Schweder, T., Hjort, N. L. (1996). Bayesian synthesis or likelihood synthesis – What does Borel's paradox say? *Report of the International Whaling Commission*, 46, 475–479.

[9] Proschan, M. A., Presnell, B. (1998). Expect the unexpected from conditional expectation. *The American Statistician*, 52 (3), 248–252.

[10] Poole, D., Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, 95 (452), 1244–1255.

[11] Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

[12] Grientschnig, D., Lira, I. (2014). Combining probability distributions by multiplication in metrology: A viable method? *International Statistical Review*, 82 (3), 392–410.

[13] Lira, I., Grientschnig, D. (2014). Deriving PDFs for interrelated quantities: What to do if there is 'more than enough' information? *IEEE Transactions on Instrumentation and Measurement*, 63 (8), 1937–1946.

[14] Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

[15] Grientschnig, D., Lira, I. (2011). Reassessment of a calibration model by Bayesian reference analysis. *Metrologia*, 48 (1), L7–L11.

[16] Grientschnig, D., Lira, I. (2012). Revision of 'Reassessment of a calibration model by Bayesian reference analysis'. *Metrologia*, 49 (1), L1–L3.

[17] Niederreiter, H. G. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84 (6), 957–1041.

[18] Berntsen, J., Espelid, T. O., Genz, A. (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software*, 17 (4), 437–451.

[19] Sun, D., Berger, J. O. (1998). Reference priors with partial information. *Biometrika*, 85 (1), 55–71.

[20] Lira, I., Grientschnig, D. (2015). The cosine error: An example of a model with a non-repetitive systematic effect. In *XXI IMEKO World Congress*, Prague, Czech Republic. `https://www.imeko.org/publications/wc-2015/IMEKO-WC-2015-TC21-404.pdf`.

[21] Datta, G. S., Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics*, 24 (1), 141–159.

[22] Datta, G. S., Ghosh, M. (1995). Some remarks on noninformative priors. *Journal of the American Statistical Association*, 90 (432), 1357–1363.

[23] Gutiérrez-Peña, E., Rueda, R. (2003). Reference priors for exponential families. *Journal of Statistical Planning and Inference*, 110 (1–2), 35–54.

[24] Berger, J. O., Bernardo, J. M. (1992). Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, Springer, 177–194.

Received February 24, 2016.
Accepted August 1, 2016.